UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/670,715	09/25/2003	David A. Luick	ROC920030293US1	6101

46797   7590   04/02/2009
IBM CORPORATION, INTELLECTUAL PROPERTY LAW
DEPT 917, BLDG. 006-1
3605 HIGHWAY 52 NORTH
ROCHESTER, MN 55901-7829

EXAMINER
ROJAS, MIDYS

ART UNIT	PAPER NUMBER
2185	

MAIL DATE	DELIVERY MODE
04/02/2009	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

	Application No.	Applicant(s)
Office Action Summary	10/670,715	LUICK, DAVID A.
	Examiner	Art Unit	
	MIDYS ROJAS	2185	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) OR THIRTY (30) DAYS,
WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed
  after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
  Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any
  earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *16 January 2009*.

2a)☐ This action is **FINAL.**   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is
  closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-12* is/are pending in the application.

  4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-12* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *25 September 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

  Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

  Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

  a)☐ All   b)☐ Some *   c)☐ None of:

    1.☐ Certified copies of the priority documents have been received.

    2.☐ Certified copies of the priority documents have been received in Application No. _____.

    3.☐ Copies of the certified copies of the priority documents have been received in this National Stage
      application from the International Bureau (PCT Rule 17.2(a)).

  * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☐ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
  Paper No(s)/Mail Date _____.

4)☐ Interview Summary (PTO-413)
  Paper No(s)/Mail Date. _____.

5)☐ Notice of Informal Patent Application (PTO-152)

6)☐ Other: _____.

DETAILED ACTION

A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 1/16/2009 has been entered.

Response to Arguments

1.   Applicant's arguments with respect to claims 1-12 have been considered but are not persuasive.

Applicant argues that the cited references do not teach that a latency to access the allocated lines of the first private cache by the second processor is greater than a latency to access cache lines of a second private cache associated with the second processor, wherein the first private cache and the second private cache are at a same cache level.

The examiner disagrees.

Dean teaches that another architecture that can be implemented for a multiple processor system is one where each processor has its own private cache (Col. 2, lines 46-52). In this implementation wherein each processor has its own private cache, each private cache is at the same cache level in relation to its associated processor. For example, cache 1 is private to processor 1 and cache 2 is private to processor 2

wherein cache 1 and cache 2 are both level 1 caches of processor 1 and processor 2 respectively. Therefore, Dean teaches that "the first private cache and the second private cache are at a same cache level".

Dean et al. does not specify that the latency to access the allocated lines of the first private cache by the second processor is greater that a latency to access cache lines of a second private cache associated with the second processor. Hetherington et al. discloses that accessing an on-chip cache provides the lowest access latency while accessing an external cache provides higher access latency (paragraph 0007). Even though Hetherington is referring to caches of different cache levels, Hetherington's disclosure still applies to the teachings of Dean. Since the limitations are drawn to a second processor accessing lines of a first private cache of a first processor in comparison to that second processor accessing lines of its second private cache; when the processor is accessing the first private cache, it is accessing an external cache in relation to itself while when the processor is access its second private cache, it is accessing its local level 1 cache.

It would have been obvious to one of ordinary skill in the art at the time the invention was made that the system of Dean et al. would provide for a higher access latency to access the allocated lines of the first private cache by the second processor and a lower access latency to access cache lines of a second private cache associated with the second processors since based on the second teaching of Dean, the resulting invention would be allocating cache ways from an external cache (cache associated with processor B) to processor A. Following the teaching of Hetherington et al., the

access latency for accessing the cache ways that are being allocated to processor A

form processor B's cache are higher than the access latency for accessing the cache

that is associated to processor A. Since the cache that is private to processor A is a

local cache, it provides the lowest latency. In allowing processor A to allocate cache

ways that are originally associated to processor B, processor A is accessing an external

cache, thus representing a higher latency.

Claim Rejections - 35 USC § 103

2. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set
> forth in section 102 of this title, if the differences between the subject matter sought to be patented and
> the prior art are such that the subject matter as a whole would have been obvious at the time the
> invention was made to a person having ordinary skill in the art to which said subject matter pertains.
> Patentability shall not be negatived by the manner in which the invention was made.

3. Claims 1-12 are rejected under 35 U.S.C. 103(a) as being unpatentable over Dean

et al. (US 6,604,174) in view of Hetherington et al. (US 2001/0010069 A1).

Regarding Claim 1, Dean et al. discloses a method for reducing latencies

associated with accessing memory for more than one processors (Proc1 110 to ProcM

112, Figure 1, wherein "processes 110, 111, and 112 can be individual processors...",

Col. 4, lines 15-16), each coupled with an associated cache 130, the method

comprising:

determining cache miss rates of the more than one processors (cache miss

percentage 194, Col. 4, lines 48-58) when issuing cache requests against the caches

(hit/miss indications 190 are used to determine the cache miss percentage);

comparing the cache miss rates of the more than one processors (each cache miss counter for each processor in system metric 191 is compared to the others, Col. 9, lines 49-62); and

allocating cache lines from a first private cache associated with a first processor to a second processor based upon the difference between the cache miss rate for the first processor and the cache miss rates of the second processor ("if a processor A's miss counter is larger than processor B's miss counter by a predetermined cache reallocation factor, some ways of the cache will be assigned to processor A", wherein cache ways comprise cache lines, Col. 9, lines 49-63).

Dean's invention does not teach the use of a private cache for each processor. Instead, Dean's invention uses a single unified cache wherein groups of cache ways are allocated to each processor so that each group of cache ways acts as a private cache section for that processor. However, Dean also teaches that another architecture that can be implemented for a multiple processor system is one where each processor has its own private cache (Col. 2, lines 46-52).   In this implementation wherein each processor has its own private cache, each private cache is at the same cache level in relation to its associated processor. For example, cache 1 is private to processor 1 and cache 2 is private to processor 2 wherein cache 1 and cache 2 are both level 1 caches of processor 1 and processor 2 respectively. Therefore, Dean teaches that "the first private cache and the second private cache are at a same cache level".

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the invention of Dean to implement the private cache

architecture (also disclosed by Dean) since the groups of cache ways disclosed by Dean already act as independent caches and independent caches also contain cache ways. Therefore, the cache way allocation techniques of the invention could be well implemented in a private cache system. Additionally, it is well known in the art that independent caches provide for faster processor access, thus improving performance and reducing latency.

Dean et al. does not teach that the latency to access the allocated lines of the first private cache by the second processor is greater that a latency to access cache lines of a second private cache associated with the second processor. Hetherington et al. discloses that accessing an on-chip cache provides the lowest access latency while accessing an external cache provides a higher access latency (paragraph 0007).

It would have been obvious to one of ordinary skill in the art at the time the invention was made that the system of Dean et al. would provide for a higher access latency to access the allocated lines of the first private cache by the second processor and a lower access latency to access cache lines of a second private cache associated with the second processors since based on the second teaching of Dean, the resulting invention would be allocating cache ways from an external cache (cache associated with processor B) to processor A. Following the teaching of Hetherington et al., the access latency for accessing the cache ways that are being allocated to processor A form processor B's cache are higher than the access latency for accessing the cache that is associated to processor A. Since the cache that is private to processor A is a local cache, it provides the lowest latency. In allowing processor A to allocate cache

ways that are originally associated to processor B, processor A is accessing an external cache, thus representing a higher latency.

Claim 5 is rejected using the same rationale as that of Claim 1 wherein the threshold cache miss rate is represented by the predetermined cache reallocation factor 195 (Col. 10, lines 4-10). Additionally, in reallocating the cache ways, cache requests associated with the first processor (processor A) will be forwarded to the way that was previously owned by the second processor (reallocated way of processor B). The cache lines in the reallocated way will be replaced with those needed by processor A (see Col. 11, line 29 – Col. 12, line 7).

Regarding Claim 2, Dean et al. in view of Hetherington et al. discloses the method wherein determining the cache miss rates comprises counting cache misses of each of the more than one processors (hit/miss indications 190 or historical files, Col. 4, lines 24-30).

Regarding Claim 3, Dean al. in view of Hetherington et al. discloses the method wherein allocating cache lines comprises forwarding cache requests from the processor to a private cache associated with another processor. In reallocating the cache ways, cache requests associated with the first processor (processor A) will be forwarded to the way that was previously owned by the second processor (reallocated way of processor B). The cache lines in the reallocated way will be replaced with those needed by processor A (see Col. 11, line 29 – Col. 12, line 7).

Regarding Claim 4, Dean al. in view of Hetherington et al. discloses the method wherein allocating cache lines comprises selectively allocating cache lines based upon

a priority associated with a cache request of the processor (allocation of cache ways, wherein cache ways have many cache lines, is based on the cache miss percentage wherein the processor with the highest cache miss percentage is given priority and assigned new cache ways first, Col. 10, lines 19-40).

Claim 6 is rejected using the same rationale as that of Claim 2 wherein the counting of the cache misses starts as soon as the system boots (since all cache accesses are taken into account when counting the total number of misses) therefore, this must occur after a cold start and warm-up period.

Regarding Claim 7, Dean al. in view of Hetherington et al. discloses the method wherein comparing the cache miss rates comprises comparing the cache miss rates, the cache miss rates being associated with more than one processor modules (each cache miss counter for each processor is compared to the others, Col. 9, lines 49-62).

Regarding Claim 8, Dean al. in view of Hetherington et al. discloses the method wherein the threshold cache miss rate predetermined cache reallocation factor is based upon an average cache miss rate for the more than one processors (see Col. 10, lines 4-10 and Col. 4, lines 48-58).

Regarding Claims 9-10, Dean al. in view of Hetherington et al. discloses the method wherein forwarding the cache request comprises selecting the second private cache based upon a least recently used cache line associated with the private caches (allocation of cache ways, wherein cache ways have many cache lines, is based on the cache miss percentage wherein the processor with the highest cache miss percentage is given priority and assigned new cache ways first, Col. 10, lines 19-40. This means

that the processor with a least recently used way, due to a low cache miss percentage, gives up a cache way to allocate it to the processor with the high miss percentage).

Regarding Claim 11, Dean al. in view of Hetherington et al. discloses the method wherein forwarding the cache request comprises selecting the cache request based upon a priority associated with the cache request (LRU algorithm preferentially writes over a process' data when that data is in a way assigned to a different process, Col. 11, lines 55-67).   The LRU information 740 is representative of the least recently cache line table.

Regarding Claim 12, Dean al. in view of Hetherington et al. discloses the method wherein forwarding the cache request is responsive to a software instruction that overrides a result of comparing the cache miss rates to forward the cache request to the second private cache (updating of allocation way assignment performed by tag allocation controller 161, see Col. 11, lines 40-55).

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to MIDYS ROJAS whose telephone number is (571)272-4207. The examiner can normally be reached on M-TH 6:00am - 4:30pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Sanjiv Shah can be reached on (571) 272-4098. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Sanjiv Shah/   /Midys Rojas/
Supervisory Patent Examiner, Art Unit 2185   Examiner, Art Unit 2185
/MR/